

VU Research Portal

Investigating Group Differences using Spearman's Hypothesis: An Evaluation of Jensen's Method

Lubke, G.H.; Dolan, C.V.; Kelderman, H.

published in

Multivariate Behavioral Research
2001

DOI (link to publisher)

[10.1207/S15327906299-324](https://doi.org/10.1207/S15327906299-324)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Lubke, G. H., Dolan, C. V., & Kelderman, H. (2001). Investigating Group Differences using Spearman's Hypothesis: An Evaluation of Jensen's Method. *Multivariate Behavioral Research*, 36(3), 299-324.
<https://doi.org/10.1207/S15327906299-324>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Investigating Group Differences on Cognitive Tests Using Spearman's Hypothesis: An Evaluation of Jensen's Method

G. H. Lubke

Free University Amsterdam

C. V. Dolan

University of Amsterdam

H. Kelderman

Free University Amsterdam

Jensen has posited a research method to investigate group differences in cognitive tests. This method consists of first extracting a general intelligence factor by means of exploratory factor analysis. Secondly, similarity of factor loadings across groups is evaluated in an attempt to ensure that the same constructs are measured. Finally, the correlation is computed between the loadings of the tests on the general intelligence factor and the mean differences between groups on the tests. This part is referred to as a test of "Spearman's Hypothesis", which essentially states that differences in g account for the main part of differences in observed scores. Based on the correlation, inferences are made with respect to group differences in general intelligence.

The validity of these inferences is investigated and compared to the validity of inferences based on multi-group confirmatory factor analysis. For this comparison, population covariance matrices are constructed which incorporate violations of the central assumption underlying Jensen's method concerning the existence of g and/or violations of Spearman's Hypothesis. It is demonstrated that Jensen's method is quite insensitive to the violations. This lack of specificity is observed consistently for all types of violations introduced in the present study. Multi-group confirmatory factor analysis emerges as clearly superior to Jensen's method.

Introduction

Differences between blacks and whites in the US on current cognitive tests have been studied extensively (Jensen, 1985, 1997; Jensen & Reynolds, 1982). It is well established that blacks, on average, score lower on a variety

The research of Conor Dolan was made possible by a fellowship of the Royal Netherlands Academy of the Arts and Sciences. We wish to thank G. J. Mellenbergh and H. van der Flier for their valuable comments on earlier drafts of this article.

Correspondence concerning this article should be addressed to Gitta H. Lubke, UCLA, Graduate School of Education & Information Studies, Moore Hall, Box 951521, Los Angeles, CA 90095-1521, or email to glubke@ucla.edu.

of psychometric tests measuring cognitive abilities (Jensen, 1985, 1998). These group differences raise two related differential psychological issues, namely how exactly do blacks and whites differ, and, secondly, why do they differ. Clearly, the former issue has to be resolved before the latter can be addressed. In an attempt to address the question of how blacks and whites differ, Jensen has posited a research method based on factor analysis (Jensen, 1985, 1992; Jensen & Reynolds, 1982). This method is designed mainly to investigate the hypothesis that black-white differences are due predominantly to a difference in general intelligence, or g , as it is denoted.

It is striking that Jensen's method is widely appreciated (for a brief overview, see Schönemann, 1997a, pp. 666-667) and used (Jensen, 1985; Lynn, 1994; Te Nijenhuis & Van der Flier, 1997; Rushton, 1999), even though the method has not been validated. Jensen's method has been subject to serious criticism (see, for example, special issues of *Multivariate Behavioral Research*, 1992, & *Cahiers de Psychologie Cognitive*, 1997) and confirmatory factor analysis has been advocated as a more adequate method to investigate group differences (Gustafsson, 1992; Millsap, 1997). However, no attempt has been made to compare the validity of the two methods. The objective of the present study is to focus on one aspect of validity, namely whether Jensen's method and multi-group confirmatory factor analysis (MGCFA) lead to rejection of the hypothesis that g is central to group differences when data are incompatible with this hypothesis. Before describing our approach in more detail, we give a short overview of the essential features of Jensen's method.

Jensen's method involves the following three steps. First, exploratory factor analytical methods such as principal component analysis (PCA), principal factor analysis (PFA), or Schmid-Leiman hierarchical factor analysis (SL-HFA) are applied (Schmid & Leiman, 1957) to cognitive test data (e.g., WISC, K-ABC) of representative black and white samples. The first principal component, the first principal factor, or, in the case of SL-HFA, the highest order factor is interpreted as the "general intelligence factor", g (Jensen, 1985, 1997; Jensen & Reynolds, 1982; Naglieri & Jensen, 1987). As a consequence of using exploratory methods, the existence of a single general intelligence factor g is usually not tested against alternative factor models with different factor structures.

The second step of Jensen's method concerns the question whether the tests measure the same constructs in both groups, in other words, whether the tests are measurement invariant. To address this question, Jensen computes a measure of congruence of the factor loadings in the white and the black samples (Jensen, 1985). However, even invariance of factor loadings across groups would not be sufficient to establish measurement invariance. A test is

said to be measurement invariant with respect to group membership if the distribution of test scores depends only on the underlying factor(s) and not on group membership (Mellenbergh, 1989; Meredith, 1993). A test is weakly measurement invariant if the expected value and the variance of test scores conditional on the underlying factor are independent of group membership (Meredith, 1993). Now suppose that the slope of the regression of observed scores on the underlying factor is equal across groups (i.e., equal factor loadings) but that the intercept of the regression differs. Intercept differences are inconsistent with weak measurement invariance because in that case the expected test score given a certain level of ability differs across groups. Differences in residual variance equally introduce bias. If, for instance, admission decisions are based on test results, the number of false admissions and false rejections given a required level of ability will be higher in the group with the larger residuals. The composite restriction of equality of factor loadings, intercepts, and residual variances of the regression of observed scores on the factor scores is termed 'strict factorial invariance' (Meredith, 1993). Strict factorial invariance is a preliminary to weak measurement invariance (Meredith, 1993, p. 538), and, consequently, to compare groups in a meaningful way. Since the necessary conditions for meaningful group comparisons within the context of factor analysis have been thoroughly discussed elsewhere (Bloxom, 1972; Ellis, 1993; Meredith, 1964, 1993), we do not evaluate the second step of Jensen's method (i.e., his partial test of measurement invariance). Instead, we focus on the validity of Jensen's method for data which are strictly factorial invariant.

The third step of Jensen's method involves testing whether the observed group differences can be attributed to differences in g . Jensen refers to this part of the method as a test of 'Spearman's Hypothesis'. Spearman's Hypothesis states that mean differences between blacks and whites on cognitive tests are a function of the tests' g -loadings: tests with higher loadings show larger mean differences than tests with low loadings (Jensen, 1992). A high correlation between the differences in means on the tests and the tests' loadings (henceforth denoted as Spearman correlation) is regarded as evidence in support of Spearman's Hypothesis. According to Jensen, confirmation of Spearman's Hypothesis allows for the conclusion that the observed mean differences are attributable to differences in g . Jensen distinguishes between a strong and a weak version of Spearman's Hypothesis. The strong version holds that "the magnitude of the black-white differences on a variety of tests are directly related to the tests' g -loadings, because black and white populations differ only on g and on no other cognitive factor" (Jensen, 1985, p. 198). The weak version states that the observed differences are mainly a function of the tests' g -loadings. Blacks

and whites may differ with regard to specific cognitive factors (e.g., spatial, or verbal ability), although to a much lesser extent than with regard to *g*. Jensen tests the strong against the weak version in a separate procedure: factor scores are computed for all subjects for the general and specific intelligence factors. These scores are then entered in a multiple regression as predictors of the dichotomous variable 'race' to test whether the specific intelligence factors significantly improve the prediction based on *g* alone (Jensen & Reynolds, 1982).

Jensen's method is based on the assumption that *g* exists. Given similar patterns of factor loadings in the two groups, the Spearman correlation is used as evidence that *g* is of central importance to black-white differences. Although Jensen has advocated confirmatory factor analysis to "obtain a good *g*" and proposed a second order factor model with uncorrelated first order factors as a preferable *g*-model (Jensen & Weng, 1994), in practice neither the assumption concerning the existence of *g*, nor strict factorial invariance, nor the attribution of observed mean differences to differences in *g*, are tested with measures of goodness of fit. In view of the ease with which these three hypotheses can be tested with programs such as LISREL (Jöreskog & Sörbom, 1993), EQS (Bentler, 1993), or Mx (Neale, 1997), this is surprising.

The fit of a *g*-model consisting of one second order factor (i.e., *g*) and uncorrelated first order factors (i.e., the specific intelligence factors), which is the factor structure resulting from SL-HFA, can be established using multi-group confirmatory factor analysis (MGCFA). Strict factorial invariance can be tested easily using MGCFA by imposing equality constraints on the factor loadings, error variances, and intercepts. Also, the hypothesis that groups differ only with respect to *g* but not with respect to the first order factors (i.e., strong version of Spearman's Hypothesis) is readily specified using MGCFA with structured means as proposed by Sörbom (1974). Unfortunately, testing the weak version is less straightforward. Jensen has not provided a precise definition of the weak version of Spearman's Hypothesis. It is unclear exactly how many of the first order factors may contribute to the observed differences in means, or, for that matter, the minimum percentage *g* has to contribute to allow for the conclusion that observed differences in means are *predominantly* due to differences in general intelligence. If such criteria are defined, all three steps of Jensen's method can be evaluated based on measures of goodness of fit using MGCFA.

As stated above, the objective of the present study is to evaluate the validity of inferences concerning black-white differences based on Jensen's method. Dolan (2000) has re-analyzed data previously published by Jensen and Reynolds (1982) using MGCFA and showed that Jensen's conclusion

that the observed black-white differences were mainly due to g was only one of several possible conclusions. Models without a predominant general intelligence factor provided an equally good description of the data. Our approach in the present study consists of comparing Jensen's method to MGCFA with structured means based on artificial data. This allows to incorporate violations of either (a) the assumption underlying Jensen's method that g exists or (b) that differences in g account for the (main part of the) differences in observed scores or (c) both. We construct population covariance matrices and mean vectors for two groups. Analysis is carried out using both MGCFA and Jensen's method. To evaluate MGCFA, the power to reject the g -model is calculated. Evaluation of Jensen's method is based on the size of Spearman correlations. The outcomes of the analyses are compared.

The outline of the present article is as follows. First, we present the factor model used to construct the data for this study. Next, our method to investigate the adequacy of Jensen's method is discussed in more detail. The description of our manipulations of the covariance and mean structure is provided as well as a brief overview of the power analysis and Schmid-Leiman hierarchical factor analysis. The presentation of the results is followed by a discussion.

Models

Our approach consists of computing population covariance matrices and mean vectors of tests for two groups using a second order factor model with a single second order factor and four first order factors. Each of the four first order factors has 4 indicators so that we have 16 observed variables. We choose four first order factors although in analysis of real data three first order factors have been reported (Dolan, 2000; Jensen & Reynolds, 1982). This choice is motivated by the fact that a covariance structure model with three correlated first order factors can always be reparameterized as a second order factor model with a single second order factor accounting for all correlations between first order factors. In other words, a model with g cannot easily be tested against a model without g . However, if we have four first order factors, a second order factor does not necessarily account for all correlations among first order factors. The first order factor model with four correlated factors is less restrictive than a second order model with a single second order factor and four uncorrelated first order factors, which allows for testing hypotheses concerning the existence of g .

Parameters of the second order factor model are gradually varied such that either the resulting covariance structure, or the mean structure, or both

deviate increasingly from a second order g -model with predominant group differences in g . More specifically, in Case 1 we add covariances between first order factors to create models which are incompatible with Spearman's hypothesis with respect to the covariance structure. The first order factors in the second order g -model described by Jensen are uncorrelated (Jensen, 1994). Manipulating the parameters of the latent mean structure in Case 2 allows us to specify models with decreasing g -contribution to group differences. Finally, in Case 3 the violations of Spearman's hypothesis of Case 1 and Case 2 are combined. The manipulations are explained in detail in the method section.

In the remainder of this section, we specify the second order factor model with four first order factors that is used to compute the population covariance matrices and mean vectors. This is followed by an overview of the parameter values we use for the computations. Then the second order g -model is described, which is a special case of the model used to construct the data.

Second Order Factor Model with Four First Order Factors

We use a special case of the LISREL Submodel 3A with structured means to compute the covariance matrices and mean vectors (Jöreskog & Sörbom, 1993). This second order factor model can be conceptualized as follows. The random vector of observations, \mathbf{y}_{ij} , of subject j in group i are defined in terms of the common factor model:

$$(1) \quad \mathbf{y}_{ij} = \mathbf{v}_i + \mathbf{\Lambda}_i \boldsymbol{\eta}_{ij} + \boldsymbol{\varepsilon}_{ij},$$

where \mathbf{v}_i is the vector of intercepts, $\mathbf{\Lambda}_i$ is the matrix of first order factor loadings, $\boldsymbol{\eta}_{ij}$ is the factor score of subject j in group i , and $\boldsymbol{\varepsilon}_{ij}$ is the vector of residuals of observed variables, which is distributed as $\boldsymbol{\varepsilon}_{ij} \sim N(0, \boldsymbol{\Theta}_i)$. Since we have 4 indicators for each of the 4 factors, \mathbf{y} , \mathbf{v} and $\boldsymbol{\varepsilon}$ have dimension 16×1 , and $\mathbf{\Lambda}_i$ is 16×4 . The residuals of observed variables are uncorrelated (i.e., $\boldsymbol{\Theta}_i$ is diagonal). The scale of the factors is determined by fixing one factor loading in $\mathbf{\Lambda}$ equal to one for each factor. The four factors are correlated. The matrix of factor loadings contains structural zeros to ensure a unique solution of the factor structure. We denote this model as "first order factor model".

To obtain the second order factor model, the first order factor scores, $\boldsymbol{\eta}$, are in turn subjected to a factor model: a single factor model is assumed to hold for $\boldsymbol{\eta}$, such that

$$(2) \quad \boldsymbol{\eta}_{ij} = \boldsymbol{\Gamma}_i \boldsymbol{\xi}_{ij} + \boldsymbol{\zeta}_{ij}.$$

Γ_i is the matrix of factor loadings of the first order factors on the single second order factor (henceforth “second order factor loadings”). The single second order factor, ξ , is distributed as $\xi_{ij} \sim N(\kappa_i, \Phi_i)$ and the 4×1 vector of first order residuals, ζ , as $\zeta_{ij} \sim N(\alpha_i, \Psi_i)$. Correlations between first order residuals can be interpreted as that part of the common variance of first order factors that is not explained by the single second order factor. If the single second order factor fully accounts for the covariances between first order factors, the matrix of first order residuals, Ψ_i , is diagonal.

Assuming that $E[(\zeta_{ij} - \alpha_i)\mathbf{e}_{ij}^t] = E[(\xi_{ij} - \kappa_i)\mathbf{e}_{ij}^t] = E[(\xi_{ij} - \kappa_i)(\zeta_{ij} - \alpha_i)^t] = \mathbf{0}$, we have $\mathbf{y}_{ij} \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where

$$(3) \quad \boldsymbol{\mu}_i = \boldsymbol{\nu}_i + \Lambda_i \alpha_i + \Lambda_i \Gamma_i \kappa_i,$$

$$(4) \quad \boldsymbol{\Sigma}_i = \Lambda_i [\Gamma_i \Phi_i \Gamma_i^t + \Psi_i] \Lambda_i^t + \Theta_i$$

The model for the observed means, $\boldsymbol{\mu}_i$, as denoted in Equation 3 is not identified. As explained by Sörbom (1974), only group differences in latent means can be modeled, not the latent means in both groups. Estimation of the latent group differences can be accomplished by setting the latent means in one group equal to zero and $\boldsymbol{\nu}$ equal across groups. The estimated latent means in the other group then represent the latent mean differences between groups. In case of the second order factor model, however, this does not yet yield unique estimates of the differences in latent means because there remains an indeterminacy between first and second order factor means. If mean differences in second order factors, κ , are estimated, only $(p - 1)$ elements of a p -dimensional vector of mean differences in first order residuals, α , are identified (i.e., one element in α has to be fixed to zero). With these constraints, the parameters in α are interpreted as that part of the group differences on the tests which is not accounted for by the difference in the second order factor.

The group index, i , is substituted by indices b and w , which denote blacks and whites, respectively. The subject parameter j is omitted to simplify notation. We arbitrarily choose to estimate the latent mean differences in the black group. This results in:

$$(5) \quad \boldsymbol{\mu}_w = \boldsymbol{\nu},$$

$$(6) \quad \boldsymbol{\Sigma}_w = \Lambda [\Gamma \Phi_w \Gamma^t + \Psi_w] \Lambda^t + \Theta,$$

$$(7) \quad \boldsymbol{\mu}_b = \boldsymbol{\nu} + \Lambda \alpha + \Lambda \Gamma \kappa,$$

$$(8) \quad \boldsymbol{\Sigma}_b = \Lambda \Gamma \Phi_b \Gamma^t + \Psi_b] \Lambda^t + \Theta.$$

Note that first and second order factor loadings, Λ and Γ , the intercept ν , and the error variances, Θ , are identical in both groups (i.e., have no group index). In this fashion we impose strict factorial invariance (Meredith, 1993). Using Equations 5 through 8, we can specify models representing the weak version of Spearman's hypothesis. The strong version implies no differences in first order residuals in addition to differences in g , consequently Equation 7 changes to

$$(9) \quad \mu_b = \nu + \Lambda\Gamma\kappa.$$

Parameter Values

We use Equations 5 through 8 to compute population covariance matrices, Σ , and mean vectors, μ , for the two groups. Manipulations concern parameter values of the off-diagonal elements of first order residuals, Ψ_b and Ψ_w , and elements of α and κ . They are described in the method section and presented in Table 1. The remaining parameter values are not varied. Consequently, the following matrices and vectors are identical for all computations: Λ , Γ , Φ_b , Φ_w , Θ , ν . Also, the diagonal elements of Ψ_b and Ψ_w remain unchanged. The values for the variances of first and second order factors, Ψ_b , Ψ_w , Φ_b , and Φ_w are chosen similar to values found in Dolan's re-analysis of a data set published in Jensen and Reynolds (Dolan, 2000; Jensen & Reynolds, 1982). The parameter values of factor loadings, residuals and intercepts of observed variables, Λ , Γ , Θ , and ν , are chosen such that the mean differences between groups, reliabilities of the tests, and correlations between tests correspond approximately to values that have been reported by Jensen and co-workers (Jensen, 1985; Jensen & Reynolds, 1982).

The matrix of first order factor loadings, Λ , has simple structure to facilitate the understanding of the effect our manipulations on the Spearman correlation:

$$\Lambda = \begin{bmatrix} 1* & - & - & - \\ .55 & - & - & - \\ .8 & - & - & - \\ .75 & - & - & - \\ - & 1* & - & - \\ - & .65 & - & - \\ - & .5 & - & - \\ - & .7 & - & - \\ - & - & 1* & - \\ - & - & .45 & - \\ - & - & .55 & - \\ - & - & .6 & - \\ - & - & - & 1* \\ - & - & - & .6 \\ - & - & - & .75 \\ - & - & - & .7 \end{bmatrix}$$

The matrix of second order factor loadings is $\Gamma^t = [1* \ 0.67 \ 0.43 \ 0.82]$.¹ The variances of the second order factor in the two groups, Φ_B and Φ_w , are 4.6 and 3.96, respectively. The variances of first order residuals in the black and white groups are the diagonal entries of Ψ_B and Ψ_w . We have $\text{diag}(\Psi_B) = [2.3, 2.1, 2.0, 2.0]$ and $\text{diag}(\Psi_w) = [2.5, 2.2, 2.3, 2.1]$. The diagonal entries of the covariance matrix of residuals of observed variables, Θ , all equal 2. The residuals of observed variables are uncorrelated. Finally, the intercepts of the observed means, ν , all equal 10.

G-model

Although Jensen has proposed several *g*-models (Jensen, 1994), we limit the main analysis to the second order *g*-model because this model has been favored by Jensen in several articles (Jensen, 1982, 1994). Analyses of other *g*-models are presented as additional results. In terms of the LISREL model described by Equations 5 through 9, the second order *g*-model results if the

¹ When fitting models to covariance matrices constructed with these first and second order factor loadings, the 1*s are fixed to equal one. This allows estimation of factor variances.

matrix of first order residuals, Ψ , is restricted to be diagonal. The second order g -model is depicted in Figure 1 and can be regarded as the confirmatory counterpart of the exploratory second order model resulting from SL-HFA (see Method section). The strong version of Spearman's Hypothesis implies that the groups differ only with respect to the mean of the second order factor, κ . The weak version is accommodated by introducing differences in first order residuals, α , in addition to a difference in κ .

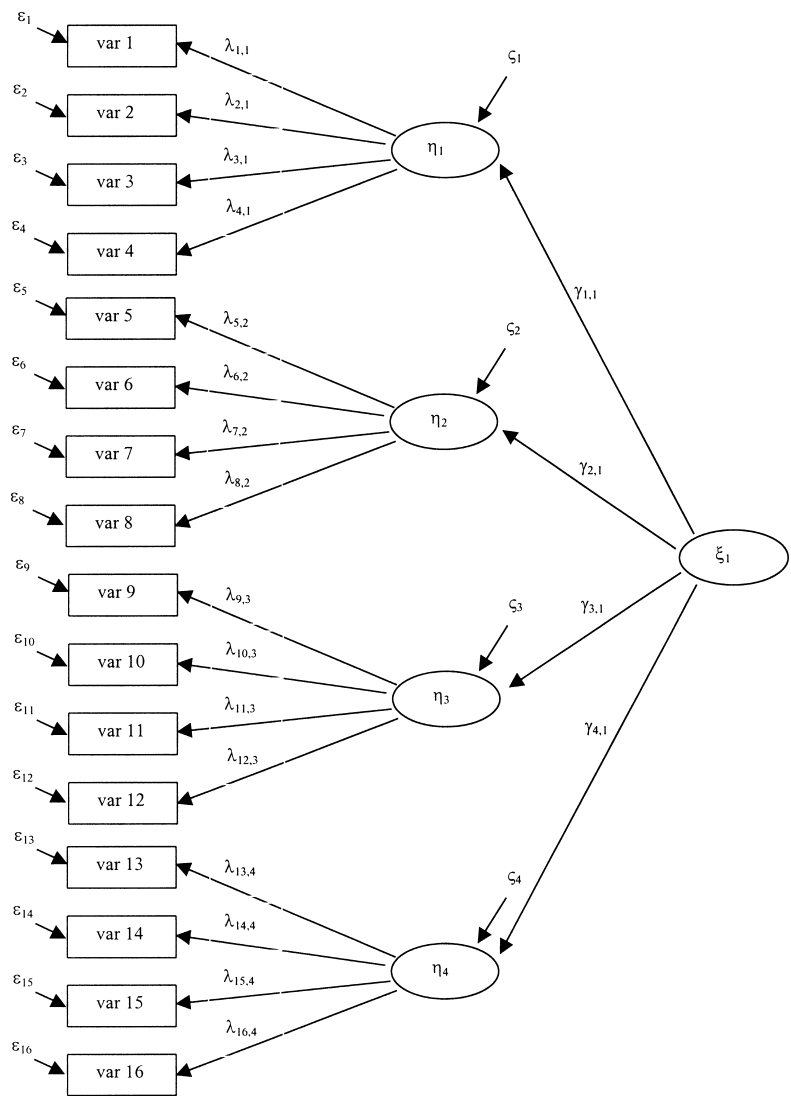


Figure 1
Path Model of the Second Order g -model

Method

The approach of this study consists of comparing inferences based on Jensen's method to inferences based on MGCFA when analyzing covariance matrices which contain increasing violations of the *g*-model with a predominant group difference in *g*. The computed covariance matrices are analyzed using SL-HFA to derive the Spearman correlation, that is, the correlation between observed mean differences and *g*-loadings. MGCFA is applied by fitting the second order *g*-model in LISREL 8.20 (Jöreskog & Sörbom, 1993). To specify the strong version of Spearman's hypothesis, we use Equations 5, 6, 9, and 8 whereas for the weak version, Equation 9 is replaced by 7.

For the comparison of inferences resulting from Jensen's method and MGCFA, we require clear rejection criteria. To evaluate the sensitivity of MGCFA to detect the violations of the *g*-model, we calculate the power to reject the *g*-model. In addition, we present the largest modification index resulting from estimating the *g*-model in LISREL. Modification indices are used as measures of misspecification. A modification index indicates the decrease in χ^2 if a constrained parameter is freed and the model is re-estimated (Sörbom, 1989). With respect to Jensen's method we investigate whether the size of the Spearman correlation results in rejection of Spearman's Hypothesis. Unfortunately, there is some uncertainty concerning the critical size of the correlation between mean differences on tests and the tests' *g*-loadings. To our knowledge, neither Jensen, nor any user of his method, has specified a lower bound for the size of the correlation beneath which one should consider Spearman's Hypothesis as disconfirmed. We therefore use the mean value as reported by Jensen (Jensen, 1985) as an indication for the size of the correlation. The mean correlation (Sd) computed for 11 studies was 0.59 (0.12). For three of the 11 studies, values between 0.3 and 0.4 were reported. Thus, it seems safe to assume that a correlation of at least 0.4 or larger would still be interpreted by users of Jensen's method as supportive of Spearman's Hypothesis.

The same uncertainty exists with regard to the contribution of *g* to observed mean differences. As stated in the introduction, the weak version of Spearman's Hypothesis is ill-defined: it is not clear what the contribution of *g* to observed mean differences has to be at least to conclude that the observed differences are "mainly" due to *g*. Jensen has reported that, although all factors contribute, *g* contributes "more than seven times as much...as the other factors combined" (Jensen & Reynolds, 1982). This equals a *g*-contribution of 87.5%, which may serve as a rough guideline when investigating whether the size of Jensen's correlation results in rejection of Spearman's Hypothesis given *g*-contributions that are clearly lower than 87.5%.

For our comparison, we consider three cases. In Case 1, the deviation from the factor structure of the g -model is obtained by gradually increasing covariances between first order residuals whereas in Case 2 mean differences in first order residuals are manipulated to obtain a decreasing g -contribution. In Case 3, the manipulations of Case 1 and Case 2 are combined.

The second order factor model described in the previous section is used to compute population covariance matrices and mean vectors. In carrying out the manipulations, care is taken that correlations between tests, and (standardized) mean differences on the tests resemble the values reported by Jensen (Jensen, 1985; Jensen & Reynolds, 1982; Naglieri & Jensen, 1987). For example, the sizes of the latent mean differences are chosen such that the standardized means of observed variables differ by roughly one standard deviation. Below, the manipulations are described for the three cases separately (see also Table 1). A short overview of the power calculation and the Schmid-Leiman procedure is given at the end of this section.

Case 1: Covariance Structure

The first step of Jensen's method consists of an exploratory factor analysis in order to extract g (see section on SL-HFA below). However, whether the model incorporating a single dominant higher order factor represents an adequate description of the data (i.e., in terms of goodness of fit) is not considered explicitly. Case 1 is meant to investigate the results of Jensen's method when the assumption of a dominant general factor is violated. It is important to assess the effect of this violation, because a researcher, upon observing a high Spearman correlation, might be tempted to construe this result as a confirmation of the presence of g . If in fact the covariance structure implied by the g -model does not fit the observed covariance structure well, such a construal would be in error.

Taking the second order factor model with a single second order factor as the initial model, positive covariances are introduced between the first and third first order residual, $\Psi_{1,3}$, and between the third and fourth first order residual, $\Psi_{3,4}$. These covariances are simultaneously increased in both groups. They are assigned the values 0.4, 0.8, and 1.2 in Cases 1a, 1b, and 1c, respectively.² As measure of deviance from Jensen's g -model we use the partial correlation between specific factors with the second order factor partialled out. The partial correlations of the first and the third first order factor and the third and the fourth first order factor equal

² Introducing higher covariances in addition to the covariance due to g is impossible since the covariance matrix of first order residuals has to remain positive definite.

approximately 0.2, 0.4 and 0.6 in Case 1a, 1b and 1c, respectively. These partial correlations can be compared to the corresponding correlations with the second order factor not partialled out, which equal 0.5, 0.6, and 0.7, respectively. Clearly, in Case 1c, the second order factor of the initial model does not play a dominant role in explaining what first order intelligence factors have in common. Case 1 can be regarded as an investigation of the validity of the first step of Jensen's method. The issue here is whether, as a consequence of using exploratory factor analysis (without explicit goodness of fit testing), application of Jensen's method gives rise to the conclusion that observed mean differences are due to differences in a single general intelligence factor when in fact there is no such dominant factor.

The mean differences on the tests in Case 1 are computed in a manner consistent with the strong version of Spearman's Hypothesis. The groups differ only with respect to the mean of the second order factor: in the black group $\kappa = -2.4$. The vector of means of first order residuals has zero entries in both groups (see Table 1). Since the violation of the g -model in Case 1 is limited to the covariance structure, in the LISREL analyses only the model representing the strong version is fitted.

Table 1
Covariance and Latent Mean Structure of Case 1a-c and Case 2a-c as Compared to the Second Order g -model Representing the Strong Version of Spearman's Hypothesis (Sp. H.)

	Covariance Matrix First Order Residuals (Ψ)	Difference in Means First Order Residuals (α)	Difference in Means Second Order Factor (κ)
g -model	$\Psi = \text{diagonal}$	$\alpha = \text{zero vector}$	$\kappa \neq 0$
Case 1a	$\Psi_{31} = 0.4$ $\Psi_{43} = 0.4$	$\alpha = \text{zero vector}$	$\kappa = -2.4$
Case 1b	$\Psi_{31} = 0.8$ $\Psi_{43} = 0.8$	$\alpha = \text{zero vector}$	$\kappa = -2.4$
Case 1c	$\Psi_{31} = 1.2$ $\Psi_{43} = 1.2$	$\alpha = \text{zero vector}$	$\kappa = -2.4$
Case 2a	$\Psi = \text{diagonal}$	$\alpha = [-2.71 -2.33 -1.3 -2.81]^t$	$\kappa = 0$
Case 2b	$\Psi = \text{diagonal}$	$\alpha = [-2.31 -2.56 -1.28 -3.05]^t$	$\kappa = 0$
Case 2c	$\Psi = \text{diagonal}$	$\alpha = [-1.9 -2.8 -1.25 -3.3]^t$	$\kappa = 0$

Case 2: Latent Mean Structure

In Case 2, the covariance structure is consistent with the g -model but the mean structure is not. Here we allow the first order residuals to contribute increasingly to the group differences on the tests. How this can be achieved becomes apparent in the following example. Suppose the vector of first order residuals, α , equals $(1\ 2\ 3\ 4)'$, the parameter of mean differences in g , κ , equals zero, and the vector of second order factor loadings, Γ , is proportional to α , taking values of, say, $(.2\ .4\ .6\ .8)'$. For reasons of simplicity, suppose further that each factor has one perfect indicator (i.e., the matrix of first order factor loadings is a 4×4 identity matrix). The mean differences in the indicators are computed as $\Lambda\alpha + \Lambda\Gamma\kappa$ (compare Equations 5 and 7) and equals in our example $\Lambda\alpha = (1\ 2\ 3\ 4)'$ since κ is zero. However, the same result can be achieved by setting the mean difference in g , κ , to 5, and fixing all elements of α to zero. Thus, a situation in which α differs from zero *and* is proportional to Γ is equivalent to a situation in which α is a zero vector and κ differs from zero.

If κ differs from zero and α is a zero vector, we have the mean structure of the model representing the strong version of Spearman's hypothesis (see Equation 9). The mean structure of the strong version of Spearman's hypothesis is therefore equivalent to the mean structure of a model without differences in g *given* proportionality of α and Γ . However, if the collinearity between α and Γ is decreased, we have differences in α in addition to differences in κ (try the example with $\Gamma = [.4\ .2\ .6\ .8]$). Setting κ to zero, the g -contribution, which is 100% in case of proportionality of α and Γ , can be decreased simply by decreasing the collinearity between α and Γ .

Recall that the vector of second order factor loadings, Γ , equals $[1\ 0.67\ 0.43\ 0.82]'$, therefore a mean vector of first order residuals, α , with values $[-3\ -2.02\ -1.28\ -2.46]$ is proportional to Γ . Although κ equals zero, this model cannot be distinguished empirically from the strong version of Spearman's hypothesis. In Case 2a, we decrease the collinearity between α and Γ by setting α to $[-2.71\ -2.33\ -1.3\ -2.81]$. This situation is equivalent to setting κ to 2.75 and α to $[0\ 0.46\ 0\ 0.54]$, meaning that both scenarios result in the same vector of observed mean differences. For Case 2b and 2c, see Table 1.

As a measure of deviation from Spearman's hypothesis, we compute the contribution of g to the observed mean differences. The g -contribution can be computed as follows. As stated above, the observed mean differences between the two groups equal $\Lambda\alpha + \Lambda\Gamma\kappa$. The g -contribution is the proportion of observed mean differences due to $\Lambda\Gamma\kappa$. Both the vector of observed mean differences, $\Lambda\alpha + \Lambda\Gamma\kappa$, and the proportion due to $\Lambda\Gamma\kappa$ can be computed using the corresponding parameter values, which are provided

for Λ and Γ in the method section and for α above. In Case 2a the mean contribution of g to the differences in the 16 indicators is 79.42%, which corresponds to the weak version of Spearman's hypothesis. Setting α to $[-2.31 \ -2.56 \ -1.28 \ -3.05]$ and $[-1.9 \ -2.8 \ -1.25 \ -3.3]$ in Case 2b and Case 2c, respectively, further decreases the collinearity of α and Γ and, consequently, the g -contribution. In Case 2b the g -contribution to observed mean differences equals 56.71% and in Case 2c 41.65%. Here, the observed mean differences are obviously not "mainly due to differences in g ".

The covariance structure of first order factors in Case 2 is in accordance to the g -model in all three subcases (see Table 1). Consequently, the covariance structure of Case 2a-2c can be adequately specified with MGCFA using the second order g -model. We first fit the model representing the strong version of Spearman's hypothesis. We allow for differences in means of first order residuals if the necessity of doing so is indicated by the modification indices (i.e., we replace Equation 9 by 7 to specify the weak version). Since differences in latent means can be accompanied by differences in variances, we allow mean differences in the first order residuals, α , to be accompanied by differences in the variances of the corresponding first order residuals. Recall that only three elements in α can be estimated in addition to the parameter representing differences in g .

Case 3: Covariance and Mean Structure

We combine each of the three subcases of Case 1 with each of the three subcases of Case 2, thereby creating composite violations of the covariance and the latent mean structure.

Power Analysis

The computed population covariance matrices and mean vectors are analyzed with LISREL 8.2. The power to reject a model is calculated as described by Saris and Satorra (1993; see also Jöreskog & Sörbom, 1993). It is important to note that the computed covariance matrices in Case 1-3 represent population covariance matrices: they do not contain sampling fluctuations. If a true model is fitted to the population covariance matrix, the resulting chi-squared goodness-of-fit index equals zero. The sampling distribution of the goodness-of-fit index under the true model is the central chi-squared distribution. The sampling χ^2 -index resulting from fitting a (not highly) misspecified model is distributed as a non-central chi-squared variate. The form of the non-central chi-squared distribution depends on the difference in degrees of freedom between the true and the misspecified

model, Δdf , the total sample size, N , and the non-centrality parameter, λ . The non-centrality parameter equals the χ^2 -index resulting from estimating the misspecified model in LISREL. Given Δdf , N , λ , and a specified significance level α (e.g., 0.05), one can calculate the power to reject the misspecified model in favor of the less restricted model. Also, given a predetermined power, one can calculate the total sample size needed to reject the misspecified model.

A requirement for the calculation of power is that the restricted model is nested under the less restricted model. However, the model representing the strong version of Spearman's hypothesis, which is fitted to all cases in this study, is not nested under the model used to compute the covariance matrices and mean vectors for Case 2. This is due to the fact that when computing the data for Case 2 we fixed the parameter for the mean difference of the second order factor, κ , to zero. Consequently, the true model is not less restrictive than the g -model representing the strong version, in which, by definition, κ differs from zero. It is not possible to compute the power to reject the 'strong' g -model in favor of the model used to compute the data. However, this problem can be solved in a rather simple manner. Both the 'strong' g -model and the second order factor model used to compute the covariance matrices of Case 1-3 are nested under a model with four correlated first order factors and differences in the means of all four factors. Fitting this model in LISREL leads to a perfect fit because the true model used to compute the data is nested under this model. We can therefore use it as the less restricted model in all Cases. We fit the g -model to the covariance matrices of Case 1-3, derive the non-centrality parameter, and compute the power to reject the g -model in favor of the model with four correlated first order factors. Choosing the model with four correlated first order factors as the less restrictive model has the interesting side effect that this model does not represent a hypothesis involving a dominant general intelligence factor. In practice, a researcher would be interested in comparing this model to the more restrictive g -model to test hypotheses concerning the existence of g .

The sample sizes in the power analysis resemble the values reported in the Jensen and Reynolds study, that is $N_b = 400$ and $N_w = 2,000$ for the black and white groups, respectively. We maintain the significance level α equaling 0.05 throughout. The number of degrees of freedom of the less restricted model equals 236. The number of degrees of freedom of the g -model depends on the number of first order residuals contributing to the differences on the tests.

Schmid-Leiman Hierarchical Factor Analysis

To investigate the effect of our manipulations on the size of the Spearman correlation, we perform Schmid-Leiman hierarchical factor analysis (Schmid & Leiman, 1957) using our own *S-Plus* routines (MathSoft, 1998). The computed covariance matrices are converted to correlation matrices which are then factor analyzed using principal factor analysis with maximum likelihood solutions. Four first order factors are extracted and these factors are then rotated using promax rotation (Lawley & Maxwell, 1971). The resulting correlation matrix of first order factors is again factor analyzed. One second order factor is extracted which represents *g*. The loadings of the tests on *g* (i.e., the *g*-loadings) are derived by multiplying the matrix of the rotated loadings of the tests on the first order factors with the matrix of loadings of the first order factors on the second order factor. Mean differences on the tests are standardized as suggested by Jensen (1985). Finally, Spearman correlations are computed by correlating the standardized mean differences with the *g*-loadings.

Next to the main analysis described above, we conducted several additional analyses including (a) the evaluation of *g*-models proposed by Jensen other than the second order *g*-model (b) the use of PCA and PFA instead of SL-HFA to compute Jensen's correlation, and (c) deviations of simple structure of the matrix of first order factor loadings, because this is usually found in practice. The outcomes of these analyses are presented as additional results.

Results

As a measure of the reliability of the indicators we use the squared multiple correlation of the LISREL output, which equals the ratio of the true variance attributable to the common first and second order factors and the total variance of an indicator. This is equivalent to computing the mean of the standardized *g*-loadings (i.e., the standardized loadings of the indicators on the second order factor, $\Lambda\Gamma$). The mean (Sd) of the squared multiple correlation of the indicators in Case 1, 2, and 3 was 0.52 (0.16). The mean of the standardized mean differences (Sd) was approximately 0.8 (0.22) in all three Cases.

Case 1

The introduction of gradually increased covariances between first order factors in Case 1a-c results in a sharp increase of power to reject the *g*-

model. The number of subjects required for an adequate power (i.e., 0.80, see Cohen, 1987) drops from 2201 in Case 1a to 227 in Case 1c. Based on a total sample size of 2400, which is approximately the sample size reported in the Jensen and Reynolds study, the power is 0.84 in Case 1a and equals 1.0 Case 1b and 1c. In addition, as can be seen in Table 3, modification indices with respect to the covariance matrix of first order residuals, Ψ , indicate that the second order *g*-model with uncorrelated first order factors contains serious misspecifications. Application of Jensen's method in Case 1a-c results in Spearman correlations (i.e., correlations between *g*-loadings and mean differences in the 16 indicators) lying between 0.78 and 0.99 (see Table 2). Clearly, these do not lead to rejection of Spearman's Hypothesis.

Case 2

Decreasing the *g*-contribution to observed mean differences from 79.42% in Case 2a through 56.71% in Case 2b, to 41.65% in Case 2c results in a drop of the Spearman correlation from 0.84 (rank order correlation) to 0.56. Given the lower bound of the Spearman correlation of 0.4, which has been chosen as a rejection criterion (see method section), we can conclude that these results would be interpreted in favor of Spearman's Hypothesis although the mean differences in Case 2c are clearly not predominantly due to *g*.

The power to reject the *g*-model representing the strong version of Spearman's Hypothesis (i.e., no differences in first order factors) with a total sample size of $N = 2400$ equals 0.84 in Case 2a and 1.0 in both Case 2b and Case 2c. The total sample size required to obtain a power of 0.80 dropped from 2,191 in Case 2a to 287 in Case 2c. In all three cases,

Table 2
Spearman Correlations Resulting from Application of Jensen's Method Based on Schmid-Leiman Hierarchical Factor Analysis (SL-HFA) for Cases 1a-c and 2a-c

	Case 1a	Case 1b	Case 1c	Case 2a	Case 2b	Case 2c
pmcc	.99	.93	.78	.92	.72	.51
rho	.99	.94	.83	.84	.64	.56

Note. Spearman correlations refer to the correlation between *g*-loadings and observed mean differences (see text). Pmcc and rho denote the product moment correlation coefficient and the rank correlation, respectively.

modification indices indicate the necessity to introduce differences in the means of first order residuals in addition to differences in g . In Case 2a, a g -model representing the weak version of Spearman's Hypothesis with differences in the second and fourth first order factor is accepted. In Case 2b and Case 2c, differences in three first order factors have to be specified to obtain non-significant modification indices in the mean model. The power to reject these g -models in favor of the less restrictive first order factor model with four correlated first order factors and differences in all four factors is poor: power equals 0.16 in Case 2a and drops to 0.05 in Case 2c. The corresponding total sample sizes necessary to obtain an adequate power are shown in Table 3.

Case 3

As shown in Table 4, only serious violations of the covariance structure in combination with serious violations of the latent mean structure (i.e., combinations of Case 1c with Case 2b and 2c) result in rank correlations

Table 3

LISREL Analyses of Case 1a-c and Case 2a-c, Power to Reject the Fitted Model with $N = 2400$, and Number of Subjects Required to Obtain Power Equaling 0.80

	Fitted Model	Non-centrality Parameter	Δdf	Largest MI	Power $N=2400$	N for Power=0.80
Case 1a	g_s	20.0	14	$\Psi_{32} = 10.84$.84	2,201
Case 1b	g_s	79.24	14	$\Psi_{32} = 43.67$	1.0	555
Case 1c	g_s	194.24	14	$\Psi_{32} = 101.86$	1.0	227
Case 2a	g_s	20.09	14	$\alpha_1 = 16.04$.84	2,191
	g_{w24}	2.87	10	-	.16	13,581
Case 2b	g_s	70.92	14	$\alpha_1 = 59.67$	1.0	621
	g_{w234}	0.29	8	-	.05	124,320
Case 2c	g_s	153.26	14	$\alpha_1 = 126.83$	1.0	287
	g_{w234}	0.29	8	-	.05	124,320

Note. MI and Δdf denote modification index and difference in degrees of freedom between the less restricted model and the fitted model, respectively. The fitted model g_s indicates the g -model representing the strong version of Spearman's Hypothesis, and g_{w234} the g -model representing the weak version with differences in the second, third, and fourth first order residuals additional to a difference in g .

between g -loadings and observed mean differences below 0.4. Using MGCFA, the power to reject the g -model with differences only in g is 1.0 in all combinations of Case 1a-c and Case 2a-c; the number of subjects required to obtain adequate power ranges between 1,027 in Case 1a/2a and 108 in Case 1c/2c. The power to reject models representing the weak version of Spearman's Hypothesis was equally adequate (or perfect) in all combinations of violations (see Table 4). Here, the number of subjects required to achieve power of 0.80 ranges between 2,368 in Case 1a/2a and 246 in Case 1c/2c.

Additional Results

Alternative g-models. Jensen has proposed two other alternative g -models in addition to the second order g -model: a simple factor model, and a bi-factor model). Both models are fitted to the covariance matrices and mean vectors of Case 1a-c. The single factor model is more restricted than the second order g -model and was clearly rejected in all cases (see Table 5). In the bi-factor model, the g -factor and the specific intelligence factors are uncorrelated first order factors (Gustafsson, 1992; Jensen, 1994). All variables load on the g -factor, whereas smaller subsets of the variables load on each of the specific intelligence factors. As shown by Schmid and Leiman (1957), the (larger number of) parameters of the bi-factor model are functions of the (smaller number of) parameters of the second order g -model. Consequently, the noncentrality parameters resulting from fitting the bi-factor model are very similar to those of the second order g -model (compare Table

Table 4
Spearman Correlations Resulting from Jensen's Method Based on SL-HFA for Combinations of Case 1a-c and Case 2a-c

	rho	Case 1a power	rho	Case 1b power	rho	Case 1c power
Case 2a	.80	.81 (g_{w24})	.76	1.0 (g_{w24})	.68	1.0 (g_{w234})
Case 2b	.58	.86 (g_{w24})	.52	1.0 (g_{w24})	.36	1.0 (g_{w234})
Case 2c	.49	.81 (g_{w234})	.43	1.0 (g_{w234})	.27	1.0 (g_{w24})

Note. Power to Reject the g -model is indicated between brackets (sample size is $N = 2400$). The notation of the g -model is as before (e.g., g_{w234} denotes the g -model representing the weak version of Spearman's Hypothesis with differences in the second, third, and fourth first order residual in addition to differences in g and ρ represents the rank correlation).

Table 5

Additional Results: Noncentrality Parameters Resulting from Fitting Alternative g-models to Case 1a-c

	Case 1a	Case 1b	Case 1c
1-factor model ($df = 254$)	4692.12	4536.01	4421.40
bi-factor model ($df = 239$)	22.42	81.52	196.32

3 and 5). However, since the bi-factor model is less parsimonious than the second order g-model, the latter model should be preferred above the bi-factor model.

Alternative Factor Analytical Methods. In the past, Jensen has used principal components analysis or principal factor analysis instead of SL-HFA to derive the correlation between g-loadings and observed mean differences. The correlations resulting from applying PCA and PFA to Case1a-c and Case2a-c are presented in Table 6. As can be seen, these Spearman correlations are consistent with those resulting from SL-HFA (compare Tables 6 and 2).

Deviations From Simple Structure. To investigate the effect of deviations from simple structure, zero values in the matrix of first order factor loadings, Λ , were replaced by values randomly drawn from the uniform distribution ranging between 0.01 and 0.25. Leaving all other matrices unchanged, we computed covariance matrices and mean vectors for Case 1a-c and conducted analyses using LISREL and Jensen's method as before. As

Table 6

Additional Results: Spearman Correlations Resulting from Jensen's Method Based on Principal Components Analysis (PCA) and Principal Factors Analysis (PFA) for Cases 1a-c and Case 2a-c

	Case 1a	Case 1b	Case 1c	Case 2a	Case 2b	Case 2c
PCA [pmcc (rho)]	.99(.99)	.97(.97)	.91(.93)	.92(.77)	.73(.53)	.53(.47)
PFA [pmcc (rho)]	.99(1.0)	.96(.97)	.90(.93)	.91(.77)	.71(.53)	.50(.47)

Note. pmcc and rho denote the product moment correlation coefficient and the rank correlation, respectively.

can be seen in Table 7, the same pattern of results emerged as in the main analysis of Case 1a-c (compare Table 2).

Discussion

The results of Case 1-3 clearly indicate that Jensen’s method is quite insensitive to violations of Spearman’s Hypothesis. Only when severe violations of the factor structure were combined with severe violations of the latent mean structure did the correlation between standardized observed mean differences of the tests and their *g*-loadings drop sufficiently to reject Spearman’s Hypothesis. Cases 1-3 show that a researcher blindly carrying out Jensen’s method may draw invalid conclusions with regard to differences in general intelligence between two groups. Specifically, a researcher may conclude that observed differences are (mainly) due to differences in general intelligence when in fact the factor structure consistent with the *g*-model does not fit the data (Case 1a-c), or when the differences are *not* mainly due to *g* (Case 2c), or when violations concern both the factor and the mean structure (Case 3). Stated otherwise, the validity of conclusions based on Jensen’s method is ambiguous. The failure of Jensen’s method can be compared to a diagnostic test in medicine with a serious lack of specificity: such a test produces too many false positives (i.e., an illness is diagnosed when in fact it is absent). The additional results of this study demonstrate that the lack of specificity of Jensen’s method is dependent neither on which type of exploratory factor analysis is used, nor on which of the proposed *g*-models is adopted, nor on whether or not the matrix of factor loadings displays strict simple structure.

Table 7
Additional Results: Spearman Correlations Resulting from Jensen’s Method Based on SL-HFA and Power to Reject the *g*-model with Modest Deviations from Simple Structure of the Matrix of First Order Factor Loadings

	Case 1a*	Case 1b*	Case 1b*
pmcc (rho)	.91 (.89)	.88 (.90)	.79 (.79)
Power <i>N</i> = 2400	1.0	1.0	1.0

Note: The star (*) denotes that Cases 1a*-c* are identical to Cases 1a-c except for the matrix of first order factor loadings. Pmcc and rho stand for the product moment correlation coefficient and the rank correlation, respectively.

The failure of Jensen's method to detect violations of Spearman's Hypothesis underlines the importance of measures of goodness-of-fit. In this respect, MGCFA is clearly superior to Jensen's method. Incompatibilities with Spearman's Hypothesis with respect to the factor structure results in rejection of the g -model, when Jensen's method shows a Spearman correlation close to unity (see Table 2). Similarly, decreasing the g -contribution to observed mean differences leads to rejection of the g -model representing the strong version of Spearman's Hypothesis. The power to discriminate between a first order model with four correlated factors and differences in all factors and the g -models representing the weak version of Spearman's Hypothesis was poor (see Case 2a-c). Here, the advantage of MGCFA over Jensen's method lies in the fact that a researcher will still make a valid inference: if models, which differ only with respect to the location of the latent mean differences, fit equally well, observed mean differences cannot be safely attributed to differences in g . Using MGCFA, combinations of violations of the covariance and the mean structure were detected in all cases. The power in Case 3 to reject models representing the strong or the weak version of Spearman's Hypothesis was consistently 0.80 or larger.

It is evident that the assumption of the existence of g is a crucial feature of Jensen's method. However, users of Jensen's method including Jensen himself do not test this assumption against competing hypotheses. MGCFA, on the other hand, provides the possibility to compare models with and without a general intelligence factor. Unfortunately, as demonstrated in Case 2, it might be difficult, if not impossible, to discriminate between competing models which incorporate only slight differences in the latent structure. This problem becomes especially acute if the number of first order factors is small. If, as has been consistently observed in practice (Jensen, 1982, 1985), the g -factor accounts for the correlations between only three specific intelligence factors, the corresponding first and second order factor models (i.e., models with and without g) are equivalent with respect to the covariance structure. Possible differences concern only the latent mean structure. As is demonstrated in Case 2a-c, the power to discriminate between these models is poor even with four first order factors. Consequently, conclusions with respect to the existence of g and with respect to differences in g cannot be drawn unequivocally. This problem has also been encountered by Dolan (2000) in a re-analysis of real data previously published by Jensen and Reynolds (1982) using MGCFA. Power studies are needed to determine the necessary conditions to test hypotheses concerning the existence of a higher order factor and concerning the location of latent mean differences.

The present study is limited to the evaluation of two aspects of Jensen's method: the assumption concerning the existence of g and testing the g -contribution to observed mean differences by means of Spearman correlations. We have not investigated Jensen's method of testing whether the same constructs are measured in the two groups. Jensen computes congruence measures for profiles of factor loadings in the black and the white group. However, since Jensen's method only uses partial information (i.e., factor loadings and standardized mean differences) and pays no attention to differences in intercepts or measurement errors, MGCFA should again be preferred. If significant differences in intercepts or measurement errors exist, one cannot strictly assume that the same constructs are measured in both groups (Meredith, 1993). The hypothesis of equal intercepts and measurement error variances is easily tested using MGCFA.

Jensen's method has been criticized extensively by Schönemann (e.g., special issues of *Multivariate Behavioral Research*, 1992, and *Cahiers de Psychology Cognitive*, 1997). In particular, he claims to have shown that in the context of PCA Jensen's method necessarily results in a correlation between observed mean differences and g -loadings close to unity. In other words Schönemann states that Jensen's method is fundamentally flawed. Our results demonstrate that, although the Spearman correlations are high in most of our cases, the size of the correlation is sensitive to extreme violations of the underlying g -model. Therefore, it seems that Jensen's method is not flawed in principle (see also, Braden, 1989). However, Schönemann's conclusions are similar to ours in that Jensen's test lacks specificity (e.g., results in a high correlation when there is not a single g -factor in the data) and that, therefore, the method is not satisfactory as a diagnostic tool to detect differences in general intelligence.

In summary, a primary issue is that Jensen's method is based on the assumption that g exist which is not tested against the competing hypothesis that there is no general intelligence factor. Jensen's method is a fragmentary approach consisting of several separate analyses. The main part of Jensen's method, namely the test of Spearman's Hypothesis, does not adequately detect violations of the underlying model. In addition, strict factorial invariance is not adequately tested. Therefore, it seems advisable to replace Jensen's method by testing composite hypotheses concerning factor structure, latent mean structure and strict factorial invariance in a single analysis using MGCFA.

References

- Bentler, P. (1993). *EQS structural equations program manual*. Los Angeles: BMDP Scientific Software.
- Bloxom, B. (1972). Alternative approaches to factorial invariance. *Psychometrika*, 37, 425-440.
- Braden, J. P. (1989). Fact or artifact? An empirical test of Spearman's hypothesis. *Intelligence*, 13, 144-199.
- Cohen, J. (1987). *Statistical power analysis for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dolan, C. V. (2000). Investigating Spearman's Hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, 35, 21-50.
- Ellis, J. L. (1993). Subpopulation invariance of patterns in covariance matrices. *British Journal of Mathematical and Statistical Psychology*, 46, 231-254.
- Gustafsson, J. E. (1992). The relevance of factor analysis for the study of group differences. *Multivariate Behavioral Research*, 27(2), 319-325.
- Jensen, A. R. (1985). The nature of the Black-White difference on various psychometric tests: Spearman's hypothesis. *Behavioral & Brain Sciences*, 8(2), 193-263.
- Jensen, A. R. (1992). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research*, 27(2), 225-233.
- Jensen, A. R. (1997). Adoption data and two g-related hypotheses. *Intelligence* 25(1), 1-6.
- Jensen, A. R. (1998). *The g-factor. The science of mental ability*. Westport: Praeger.
- Jensen, A. R. & Reynolds, C. R. (1982). Race, social class and ability patterns on the WISC-R. *Personality & Individual Differences*, 3(4), 423-438.
- Jensen, A. R. & Weng, L. J. (1994). What is a good g? *Intelligence*, 18(3), 231-258.
- Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.
- Lawley, D. N. & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth.
- Lynn, R. & Owen, K. (1994). Spearman's Hypothesis and test score differences between Whites, Indians, and Blacks in South Africa. *The Journal of General Psychology*, 121(1), 27-36.
- MathSoft. (1998). *S-PLUS 4 guide to statistics*. Seattle: Author.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127-143.
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29, 177-185.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Millsap, R. E. (1997). The investigation of Spearman's hypothesis and the failure to understand factor analysis. *Cahiers de Psychologie Cognitive*, 16(6), 750-757.
- Naglieri, J. A. & Jensen, A. R. (1987). Comparison of Black-White differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence*, 11(1), 21-43.
- Neale, M. C. (1997). *Mx: Statistical modeling*. Richmond: Medical College of Virginia.
- Rushton, J. P. (1999). Secular gains in IO not related to the g-factor and inbreeding depression – unlike Black-White differences: A reply to Flynn. *Personality and Individual Differences*, 26, 381-389.
- Saris, W. E. & Satorra, A. (1993). *Power evaluations in structural equation models*. Newbury Park: Sage.

- Schmid, J. & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53-61.
- Schönemann, P. H. (1997a). Famous artefacts: Spearman's hypothesis. *Cahiers de Psychologie Cognitive*, 16(6), 665-694.
- Schönemann, P. H. (1997b). The rise and fall of Spearman's hypothesis. *Cahiers de Psychologie Cognitive*, 16(6), 788-812.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.
- Sörbom, D. (1989). Model modification. *Psychometrika*, 54, 371-384.
- Te Nijenhuis, J. & Van der Flier, H. (1997). Comparability of GATB scores for immigrants and majority group members: Some dutch findings. *Journal of Applied Psychology*, 82, 675-687.

Accepted June, 2000.

Copyright of Multivariate Behavioral Research is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Copyright of Multivariate Behavioral Research is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.